

Claims

What is claimed is:

1. A method of managing data objects in a computer system, the method comprising the steps of:

- 5 maintaining a log of at least a portion of accesses to the data objects;
 determining from the maintained log at least one cluster comprised of data objects accessed at substantially similar times; and
 storing the data objects comprising the at least one cluster in close proximity to one another in a memory.

10 2. The method of claim 1, wherein an access comprises a request to one of read and write a data object.

 3. The method of claim 1, wherein the data objects comprise Web data and the log comprises at least one Web log.

15 4. The method of claim 1, wherein the determining step further comprises the steps of:

- determining a number of time periods, $c(a)$, a cluster is accessed;
 determining a number of time periods, $c(o)$, an object is accessed along with the cluster; and
 using $c(a)$ and $c(o)$ to determine whether to add the object to the cluster.

20 5. The method of claim 4, wherein the using step further comprises the step of computing a quotient, $c(o)/c(a)$, and adding the object to the cluster when $c(o)/c(a)$ is not less than a predetermined value.

6. The method of claim 1, wherein the determining step further comprises the steps of:

determining a number of time periods, $c(o)$, an object is accessed along with the cluster; and

5 adding the object to the cluster when $c(o)$ is not less than a predetermined value.

7. The method of claim 1, further comprising the steps of:

receiving a request for a data object in a cluster;

determining from the log a probability that at least one other data object in the cluster may be subsequently requested; and

10 in response to the probability being not less than a predetermined value, retrieving both the requested data object and the at least one other data object.

8. The method of claim 7, wherein the step of determining from the log a probability further comprises the steps of:

15 determining a number of time periods, $c(o)$, the object is accessed along with the cluster;

determining a number of time periods, $t(o)$, the object is accessed; and

determining the probability using $c(o)$ and $t(o)$.

9. The method of claim 8, wherein the probability determining step further comprises computing a quotient, $c(o)/t(o)$.

20

10. The method of claim 1, wherein the memory comprises a disk storage device.

11. Apparatus for managing data objects in a computer system, the apparatus comprising:

at least one processor operative to: (i) maintain a log of at least a portion of accesses to the data objects; (ii) determine from the maintained log at least one cluster comprised of data objects accessed at substantially similar times; and (ii) store the data objects comprising the at least one cluster in close proximity to one another in a data storage device; and

memory, operatively coupled to the at least one processor, for storing at least one of the log and a cluster membership identifying the at least one cluster.

12. The apparatus of claim 11, wherein an access comprises a request to one of read and write a data object.

13. The apparatus of claim 11, wherein the data objects comprise Web data and the log comprises at least one Web log.

14. The apparatus of claim 11, wherein the determining operation further comprises: (i) determining a number of time periods, $c(a)$, a cluster is accessed; (ii) determining a number of time periods, $c(o)$, an object is accessed along with the cluster; and (iii) using $c(a)$ and $c(o)$ to determine whether to add the object to the cluster.

15. The apparatus of claim 14, wherein the using operation further comprises computing a quotient, $c(o)/c(a)$, and adding the object to the cluster when $c(o)/c(a)$ is not less than a predetermined value.

16. The apparatus of claim 11, wherein the determining operation further comprises: (i) determining a number of time periods, $c(o)$, an object is accessed along with the cluster; and (ii) adding the object to the cluster when $c(o)$ is not less than a predetermined value.

17. The apparatus of claim 11, wherein the at least one processor is further operative to: (i) receive a request for a data object in a cluster; (ii) determine from the log a probability that at least one other data object in the cluster may be subsequently requested; and (iii) in response to the probability being not less than a predetermined value, retrieve both the requested data object and the at least one other data object.

18. The apparatus of claim 17, wherein the operation of determining from the log a probability further comprises: (i) determining a number of time periods, $c(o)$, the object is accessed along with the cluster; (ii) determining a number of time periods, $t(o)$, the object is accessed; and (iii) determining the probability using $c(o)$ and $t(o)$.

19. The apparatus of claim 18, wherein the probability determining operation further comprises computing a quotient, $c(o)/t(o)$.

20. The apparatus of claim 11, wherein the data storage device comprises a disk storage device.

21. In a system comprising at least one server and at least one disk storage device operatively coupled to the at least one server, apparatus for managing data objects in accordance with the at least one server and the at least one disk storage device, the apparatus comprising:

memory for storing at least one log, the log comprising information relating to at least a portion of accesses to the data objects; and

a module, operatively coupled to the log memory, and operative to cause the storing of the data objects in at least one cluster on the at least one disk storage device via the at least one server based on the at least one log.

22. The apparatus of claim 21, wherein the module is further operative to: (i) learn of a request for a data object in a cluster; (ii) determine from the log a probability that at least one other data object in the cluster may be subsequently requested; and (iii) in response to the probability being not less than a predetermined value, cause the retrieval of both the requested data object and the at least one other data object from the at least one disk storage device.

23. The apparatus of claim 21, wherein the at least one server is one of a Web server and a proxy server.

24. An article of manufacture for managing data objects in a computer system, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

maintaining a log of at least a portion of accesses to the data objects;

determining from the maintained log at least one cluster comprised of data objects accessed at substantially similar times; and

storing the data objects comprising the at least one cluster in close proximity to one another in a memory.

25. The article of claim 24, further comprising the steps of:

receiving a request for a data object in a cluster;

determining from the log a probability that at least one other data object in the cluster may be subsequently requested; and

in response to the probability being not less than a predetermined value, retrieving both the requested data object and the at least one other data object.